PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 :		(11) International Publication Number: WO 90/07000
C12N 15/54, 9/10, C12P 21/08 C12P 19/12, C12Q 1/68, 1/48	A2	(43) International Publication Date: 28 June 1990 (28.06.90)

(21) International Application Number: PCT/US89/05128

(22) International Filing Date: 16 November 1989 (16.11.89)

(30) Priority data:
283,732 13 December 1988 (13.12.88) US

(71) Applicant: LA JOLLA CANCER RESEARCH FOUNDATION [US/US]; 10901 North Torrey Pines Road, La Jolla, CA 92037 (US).

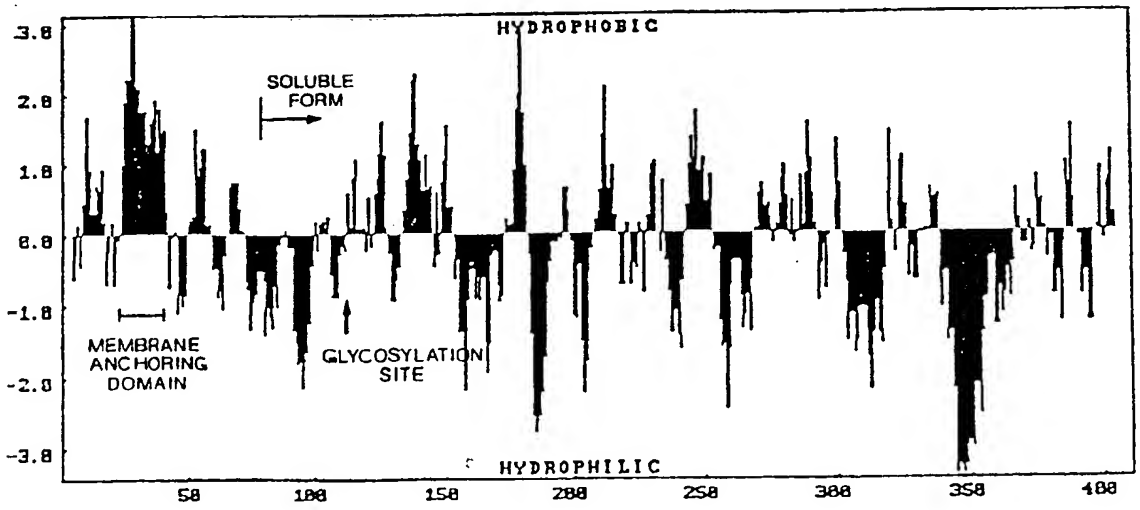
(72) Inventors: FUKUDA, Michiko, N. ; 2818 Passy Avenue, San Diego, CA 92122 (US). APPERT, Hubert, A. ; 5444 Morrow Road, Toledo, OH 43615 (US).

(74) Agents: CAMPBELL, Cathryn et al.; Pretty, Schroeder, Brueggemann & Clark, 444 South Flower Street, Suite 2000, Los Angeles, CA 90071 (US).

(81) Designated States: AT (European patent), AU, BE (European patent), CH (European patent), DE (European patent), ES (European patent), FR (European patent), GB (European patent), IT (European patent), JP, LU (European patent), NL (European patent), SE (European patent).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: NUCLEOTIDES ENCODING HUMAN B1, 4-GALACTOSYLTRANSFERASE AND USES THEREOF



(57) Abstract

The present invention provides an isolated nucleic acid sequence which encodes purified membrane-bound human β-1,4-galactosyltransferase, or a functional equivalent thereof. This invention also provides an isolated nucleic acid sequence which encodes purified soluble human β-1,4-galactosyltransferase or a functional equivalent thereof. The invention further provides vectors comprising these nucleic acid sequences and the expression of recombinant proteins by use of a host vector system. The invention still further provides antibodies reactive with the proteins and probes reactive with the nucleic acid sequences. Finally, the invention provides a method of diagnosing congenital dyserthropoietic anemia type II in a subject.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	ES	Spain	MG	Madagascar
AU	Australia	FI	Finland	ML	Mali
BB	Barbados	FR	France	MR	Mauritania
BE	Belgium	GA	Gabon	MW	Malawi
BF	Burkina Faso	GB	United Kingdom	NL	Netherlands
BG	Bulgaria	HU	Hungary	NO	Norway
BJ	Benin	IT	Italy	RO	Romania
BR	Brazil	JP	Japan	SD	Sudan
CA	Canada	KP	Democratic People's Republic of Korea	SE	Sweden
CF	Central African Republic			SN	Senegal
CG	Congo	KR	Republic of Korea	SU	Soviet Union
CH	Switzerland	LI	Liechtenstein	TD	Chad
CM	Cameroon	LK	Sri Lanka	TG	Togo
DE	Germany, Federal Republic of	LU	Luxembourg	US	United States of America
DK	Denmark	MC	Monaco		

NUCLEOTIDES ENCODING HUMAN B1, 4-GALACTOSYLTRANSFERASE AND USES THEREOF

This invention relates glycoproteins and more specifically to enzymes which catalyze glycosylation.

The subject invention was made pursuant to grant Nos. OK 37016, CA 30199 and CA 34014. The United States government may have certain rights in the invention.

BACKGROUND OF THE INVENTION

To a large extenct cells are made of proteins, which consititue more than half of the dry weight of the cell. Proteins determine the shape and structure of the cell and also serve as instruments of molecular recognition and catalysis. The biological function of a protein depends on its detailed chemical properties. A protein is often nonfunctional until it is modified in the cell. One such modification is glycoslation. Proteins which have been glycoslated are termed glycoproteins. The first step in glycosylation takes place in the endoplasmic reticulum (ER), where mainly one species of oligosaccharide is attached to proteins. Most of the differences in oligosaccharide structures found attached to different mature proteins are generated by subsequent modifications during their passage through the Golgi apparatus.

The glycosyltransferases are recognized as a functional family of intracellular, membrane-bound enzymes that participate coordinately in the biosynthesis of the carbohydrate moieties of glycoproteins and glycolipids. Specific glycosyltransferases have been demonstrated in two distinct intracellular membrane sites: the rough endoplasmic reticulum and the Golgi apparatus, where assembly of the mannose/N-acetylglucosamine core and both N-linked and O-linked glycosylation take place, respectively. The galactosyltransferases are a subset of the

glycosyltransferases that use uridine diphosphate galactose (UDP-galactose or UDP-gal) as the activated sugar donor. At least nine different galactosyltransferase activities have been described based on acceptor sugar requirements and glycosidic linkages formed.

UDP-β-1,4-galactosyltransferase (UDP-galactose:N-acetylglucosamine galactosyltransferase; EC 2.4.1.38) is widely distributed among animal tissues and catalyzes the following reaction:

$$\text{UDP-Gal} + \text{GlcNAc} \xrightarrow{Mn^{2+}} \text{GalB-1,4GlcNAc} + \text{UDP}$$

where the acceptor sugar, N-acetylglucosamine (GlcNAc), may be either the free monosaccharide or the nonreducing terminal monosaccharide of a carbohydrate side chain of a glycoprotein or glycolipid. In mammary tissue, β1,4-galactosyltransferase can also interact with the hormonally regulated protein α-lactalbumin. This complex (lactose synthetase, EC 2.4.1.22) is responsible for the biosynthesis of the unique mammalian disaccharide, lactose.

Historically, β1,4-galactosyltransferase has served as a Golgi marker enzyme for cell fractionation procedures. Subsequent immunohistochemical localization at the level of the EM has shown that the enzymes distribution is restricted to the trans-cisternae of the Golgi. β1,4-Galactosyltransferase has also been localized to the plasma membrane of a variety of cells and tissues by immunohistochemical procedures and biochemical procedures. This cell surface distribution supports the hypothesis that, in addition to its biosynthetic role, this transferase also has a functional role in intercellular recognition/adhesion.

While β1,4-galactosyltransferase is located primarily in the trans-cisternae of the Golgi complex in a membrane bound form it is also present in a soluble form in body fluids such as milk, colostrum, and serum. Pulse labeling of galactosyltransferase in cultured cells and comparison between molecular weights of the two forms suggest that the soluble form is produced from the membrane form by proteolytic cleavage. Recently, a congenital anemia patient who is defective in β1,4-galactosyltransferase among patients of congenital dyserythropoietic anemia type II (HEMPAS) has been identified (Fukuda, M.N. Masri, K.A., Dell, A., Thonar, E.J.M, Klier, G., and Lowenthal R.M., Blood, in press), incorporated by reference herein.

Appert, et al. (1986) Biochem. Biophys. Res. Comm., 139, 163-168, isolated and sequenced a cDNA coding for a portion of human β1,4-galactosyltransferase but not the N-terminal membrane-bound portion, nor the translational initiation codon. Additionally, Shaper, et al. (1988) J. Biol. Chem., 263, 10420-10428, recently identified the full-length cDNA for murine galactosyltransferase. However, a comparison of the currently available murine sequence data indicated that there was a considerable amount of amino acid sequence variation on the N-terminal part of the enzyme. Consequently, when studying human congenital defects involving β1,4-galactosyltransferase expression, sequence data obtained from non-human species would not suffice to explain whether or not the abnormality resulted from any specific DNA mutation and such data was not known for human β1,4-galactosyltransferase.

A complete nucleotide sequence of the soluble and membrane-bound form of β1,4-galactosyltransferase would allow the cloning and expression of recombinant forms of these proteins which can be used in the biosynthesis of useful

sugars, glycoproteins, or glycolipids. Additionally, the complete nucleotide sequence can be used in the production antibodies and probes for the detection of polypeptides and nucleotides, respectively, useful in the diagnosis of disorders associated with the enzymes. Thus, there exists a need which is satisfied by the present invention.

SUMMARY OF THE INVENTION

The present invention provides a isolated nucleic acid sequence which encodes purified membrane-bound human β-1,4-galactosyltransferase, or a functional equivalent thereof. This invention also provides a isolated nucleic acid sequence which encodes purified soluble human β-1,4-galactosyltransferase, or a functional equivalent thereof. The invention further provides vectors comprising the nucleic acid sequences and the expression of recombinant proteins by use of a host vector system. The invention still further provides antibodies reactive with the proteins and probes reactive with the nucleic acid sequences. Finally, the invention provides a method of diagnosing congenital dyserthropoietic anemia type II in a subject.

BRIEF DESCRIPTION OF THE DRAWINGS

FIGURE 1 shows the full length cDNA for both soluble and membrane-bound β1,4-galactosyltransferase, isolated cDNA's and sequencing strategy for presently isolated cDNA clone. A, Full length of β1,4 galactosyltransferase cDNA estimates from Northern blot analysis and characterized by full length murine galactosyltransferase cDNA. B, cDNA for Bovine galactosyltransferase encoding a partial amino acid sequence of the enzyme. C, Partial human galactosyltransferase cDNA that was used as probe for isolation of new cDNA clones. The

small box under the cDNA represents the oligonucleotide probe used in screening. D and E, Clones J20 (D) and CT7 (E) were isolated, sequenced and their combined data gives an approximately 1.4kb long sequence containing the full coding region. F, Sequencing strategy represented by arrows gives direction and length of sequence performed. The thick line represents the coding region.

FIGURE 2 shows nucleotide sequence and complete amino acid sequence of human β1,4-galactosyltransferase inferred from the nucleotide sequence of the cDNAs. Peptide sequence of the membrane anchoring signal peptide is underlined. The NH_2-terminal sequence of the purified soluble form of the enzyme is underlined with a broken line. Potential glycosylation site (Asn-X-Thr/Ser) is boxed. The $(A)_8$, where the CT7 clone is primed is highlighted.

FIGURE 3 shows a hydropathy plot of human β1,4-galactosyltransferase. Amino acid sequence was analyzed for hydrophobicity and hydrophilicity and plotted on Genepro Software (Riverside Scientific Enterprises, Seattle, WA,). Each line corresponds to one amino acid. The numbers on the bottom represent amino acid residues.

FIGURE 4 shows a comparison of β1,4-galactosyltransferase amino acid sequences between human, mouse, and bovine species. Asterisks show deletion of corresponding residues. Variation between human and mouse is 14% in the entire sequence. A comparison with human and bovine in the available area (343 residues) indicates a 16% variation.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

An isolated nucleic acid sequence which encodes purified

membrane-bound human β1,4-galactosyltransferase, or a functional equivalent thereof is provided. The nucleic acid sequence may be DNA, RNA or cDNA. An example of a cDNA sequence comprises the sequence identified for membrane-bound human β1,4-galactosyltransferase in Figure 2. The nucleic acid sequence may additionally have the sequence identified in Figure 2 beginning with adenine at position 1 and ending with cytosine at position 1200.

The invention also provides a isolated nucleic acid sequence which encodes purified soluble human β1,4-galactosyltransferase, or a functional equivalent thereof. The nucleic acid sequence may be DNA, RNA or cDNA. An example of a cDNA sequence comprises the sequence identified for soluble human β1,4-galactosyltransferase in Figure 2. The nucleic acid sequence may additionally have the sequence identified in Figure 2 beginning with adenine at position 231 and ending with cytosine at position 1200.

As used herein, "functional equivalent" means a nucleotide sequence encoding a polypeptide which has the same or a similar but improved function as β1,4-galactosyltransferase, i.e. catalyze the transfer of galactose from UDP-galactose to an acceptor sugar such as N-acetylglucosamine. Thus, minor modifications of the nucleotide sequence which improve and do not destroy the encoded enzyme activity is contemplated in the subject invention. Both forms of human β1,4-galactosyltransferase have substantially the amino acid sequence shown in Figure 2 which corresponds to the nucleotide sequence also set forth in Figure 2. Moreover, only a portion of the nucleotide sequence may be required to encode the active enzymes and this portion is within the scope of the invention.

Within the specification, "galactosyltransferase" and

"β1,4-galactosyltransferase" may be used interchangeably and are intended to refer to the same protein. Two forms of β1,4-galactosyltransferase are described herein, membrane-bound and soluble. The soluble form is produced from the membrane-bound form by proteolytic cleavage. This proteolytic cleavage occurs between arginine and threonine encoded by nucleotides 228 through 233 set forth in Figure 2. Thus, the soluble form lacks the anchoring signal peptide underlined in Figure 2. Further, the amino acid sequence for the soluble form corresponds to the sequence for the membrane-bound form beginning at theronine encoded by nucleotides 231 through 233 and ending with serine encoded by nucleotides 1198 through 1200 in Figure 2. Additionally, the functional portion of β1,4-galactosyltransferase occurs in the amino acid sequence common to the two enzyme forms.

As herein described, membrane-bound β1,4-galactosyltransferase refers to the β1,4-galactosyltransferase normally located primarily in the trans-cisternal of the Golgi complex in a membrane-bound form although the β1,4-galactosyltransferase may exist or be synthesized in a non-membrane bound form and is termed "membrane-bound" merely to distinguish it from the soluble form.

Additionally, both β1,4-galactosyltransferases, soluble and membrane-bound, may be modified by the presence of certain biological materials such as lipids and saccharides, by side chain modifications such as the acetylation of amino groups, phosphorylation of hydroxyl side groups or oxidation or reduction of sulfhydryl groups. Included within the definition of functional equivalent herein are any composition of an amino acid sequence substantially similar to that of the native human sequence. Moreover, the primary amino acid sequence may be modified, either deliberately, as

through site directed mutagenesis, or accidentally, as through mutation of host's DNA, but still retain the β1,4-galactosyltransferase activity. All such modifications including alternative splicing, are also included in the definition of functional equivalent, as long as β1,4-galactosyltransferase activity is retained.

"β1,4-galactosyltransferase activity" as used herein, denotes the ability to catalyze the transfer of galactose from UDP-galactose to acceptor sugars.

The term "nucleic acid sequence which codes for both the soluble and membrane-bound human β1,4-galactosyltransferase" as used herein refers to the primary nucleotide sequence of a gene encoding the amino acid sequence of the respective β1,4-galactosyltransferase, as defined above. An example is the sequence presented in Figure 2. The gene may or may not be expressed in the native host. If it is not expressed in the native host, it may still be capable of being manipulated through recombinant techniques to effect expression in a foreign host. The term refers both to the precise nucleotide sequence of a gene found in a mammalian host as well as modified genes which still code for polypeptides having the same or similar biological activity. The gene may exist as a single contiguous sequence or may, because of intervening sequences and the like, exist as two or more discontinuous sequences, which are nonetheless transcribed in vivo to ultimately effect the biosynthesis of a protein substantially equivalent to that defined above. Such modifications may be deliberate, resulting from, for example, site directed mutations. Such modifications may be neutral, in which case they result in redundant codons specifying the native amino acid sequence or in such modifications which may in fact result in a change in amino acid sequence which has either no effect, or only an insignificant effect on activity. Such

modifications may include point mutations, deletions or insertions.

As is well known, genes for a particular polypeptide may exist in single or multiple copies within the genome of an individual. Such duplicate genes may be identical or may have certain modifications, including nucleotide substitutions, additions or deletions, which all still code for polypeptides having substantially the same activity. The term "nucleic acid sequence coding for soluble and membrane-bound human β1,4-galactosyltransferase" may thus refer to one or more genes within a particular individual. Moreover, certain differences in nucleotide sequences may exist between individual organisms, which are called alleles. Such allelic differences may or may not result in differences in amino acid sequence of the encoded polypeptide which still encode a protein with activity.

The invention further provides a vector comprising the nucleic acid sequence of either soluble or membrane-bound β1,4-galactosyltransferase. This vector may be any known or later discovered vector including a plasmid. Examples of a suitable plasmids which may be used as vectors are pTZ18U and pIN-III-omp3.

Recombinant host cells transformed with these vectors are also provided as well as polypeptides produced by the recombinant host cells. These polypeptides include recombinant soluble and membrane-bound forms of β1,4-galactosyltransferase and their functional equivalents are defined hereinabove.

"Cells," "host cells" or "recombinant host cells" are terms used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but

to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

"Vector" includes vectors which are capable of expressing DNA sequences contained therein, where such sequences are operationally linked to other sequences capable of effecting their expression. It is implied that these expression vectors must be replicable in the host organisms either as episomes or as an integral part of the chromosomal DNA. Clearly a lack of replicability would render them effectively inoperable. In sum, "vector" is given a functional definition, and any DNA sequence which is capable of effecting expression of a specified DNA code disposed therein is included in this term as it is applied to the specified sequence. In general, vectors of utility in recombinant DNA techniques are often in the form of "plasmids" which refer to circular double stranded DNA loops which, in their vector form are not bound to the chromosome. "Plasmid" and "vector" may be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors which serve equivalent functions and which become known in the art subsequently hereto.

This invention still further provides antibodies, including monoclonal and polyclonal, reactive with a portion of membrane-bound β1,4-galactosyltransferase identified in Figure 2 beginning with arginine corresponding to nucleotide positions 4 through 6, or methionine corresponding to positions 1 through 3 in the case where methionine is part of the functional enzyme, and ending with arginine corresponding to nucleotide positions 228 through 230. This segment of

membrane-bound β1,4-galactosyltransferase represents the segment which is proteolyticly cleaved in the soluble form and is therefore unique to the membrane-bound form and may be used to distinguish the two forms.

Antibodies including monoclonal and polyclonal, reactive with a portion of both soluble and membrane-bound β1,4-galactosyltransferase identified in Figure 2 beginning with threonine corresponding to nucleotide positions 231 through 233 and ending with serine corresponding to nucleotide positions 1198 through 1200 are also provided. This segment is common to both forms of β1,4-galactosyltransferase and therefore antibodies reactive with this common portion may be used to detect both forms.

The invention also provides a nucleic acid probe comprising a nucleotide sequence complementary to a portion of the nucleotide sequence 1 to 411 in Figure 2. In a preferred embodiment the nucleotide probe is between 10 and 350 nucleotides but may be any length sufficient to hybridize with portions of the sequence characteristic of the human sequence. Such hybridization procedures are well known in the art.

Nucleic acid probes specific for a portion of nucleotides which are translated into polypeptides encoded by β1,4-galactosyltransferase can be used to detect nucleotide variation for diagnostic purposes. Nucleic acid probes suitable for such analyses can be prepared from the cloned sequences or by synthesizing oligonucleotides which hybridize only with the homologous sequence under stringent conditions. The oligonucleotides can be used as such to detect DNA, mRNA or they can be used to isolate cDNA clones from libraries. The probe can be labelled, using labels and methods well known in the art.

Antibodies to the enzyme are generated by immunizing with the enzyme or fragments thereof isolated from natural sources or produced from the cDNA in a bacterial or eukaryotic expression system by using methods well known in the art. Alternatively, antigenic peptides can be synthesized by chemical methods well known in the art. An example of an effective synthesized peptide is Ser-Arg-Asp-Lys-Lys-Asn-Glu-Pro-Asn-Pro-Gln-Arg-Phe-Asp-Arg but one skilled in the art may make a number of such peptides.

The β1,4-galactosyltransferase polypeptides can be used to produce either polyclonal or monoclonal antibodies. If polyclonal antibodies are desired, purified β1,4-galactosyltransferase proteins, or antigenic fragments thereof, which may be isolated or synthesized, are used to immunize a selected mammal (e.g. mouse, rabbit, goat, horse, etc.) and serum from the immunized animal is later collected and treated according to known procedures. The fragments may be antigenic either alone or conjugated to a carrier. Antisera containing polyclonal antibodies to a variety of antigens in addition to the desired polypeptide can be made substantially free of antibodies which are not β1,4-galactosyltransferase specific by passing the composition through a column to which non-β1,4-galactosyltransferase polypeptides prepared from the same expression system without β1,4-galactosyltransferase have been bound. After washing, antibodies to the non-β1,4-galactosyltransferase polypeptides will bind to the column, whereas anti-β1,4-galactosyltransferase antibodies elute in the flow through. Such methods are well known.

Alternatively, antisera can be purified by passing the serum through a column to which bovine galactosyltransferase (Sigma Chemical Co., St. Louis, MO) is conjugated.

Antibodies specific to galactosyltransferase can be eluted with 4M guanidine-HCl in phosphate buffered saline (PBS). The antibodies can be recovered after dialyzing out the guanidine-HCl. In order to obtain antibodies specific to a NH_2-terminal region, however, peptides conjugated to a matrix can be used for immunoabsorbent.

Monoclonal anti-β1,4-galactosyltransferase antibodies can also be readily produced by one skilled in the art. The general methodology for making monoclonal antibodies by fusing myelomas and lymphocytes to form hybridomas is well known. Such cells are screened to determine whether they secrete the desired antibodies, and can then be grown either in culture or in the peritoneal cavity of a mammal. Antibodies that can be antibody producing cell lines can also be created by techniques other than fusion, such as direct transformation of B lymphocytes with oncogenic DNA, or transfection with Epstein-Barr virus, See, e.g., M. Schreier et al., HYBRIDOMA TECHNIQUES (1980); Hammerling et al., MONOCLONAL ANTIBODIES AND T-CELL HYBRIDOMAS (1981); Kennett et al., MONOCLONAL ANTIBODIES (1980), which are incorporated herein by reference.

Antibodies specific to human β1,4-galactosyltransferase have a number of uses. For example, they may be employed in an immunoassay to detect the presence of human β1,4-galactosyltransferase or to detect a disease state associated with increased or decreased expression of the proteins. Various appropriate immunoassay formats are well known to those skilled in the art. See for example HANDBOOK OF EXPERIMENTAL IMMUNOLOGY, (D.M. Weir, Ed.) Blackwell Scientific Publications (3rd ed. 1978), which is incorporated herein by reference.

A method of catalyzing the transfer of galactose from

UDP-galactose to acceptor sugars comprising performing the transfer in the presence of β1,4-galactosyltransferase is additionally provided. The acceptor sugar may be but is not limited to N-acetylglucosamine or glucose. In the case of glucose, β1,4-galactosyltransferase interacts with α-lactalbumin and this complex is responsible for the biosynthesis of lactose from glucose.

Finally, a method of diagnosing an abnormal condition in a subject is provided. The method comprises detecting the presence of soluble and/or membrane-bound β1,4-galactosyltransferase, quantifying the relative amounts of soluble and/or membrane-bound β1,4-galactosyltransferase and comparing the amount of soluble and/or membrane-bound β1,4-galactosyltransferase to the amount in a normal subject; an increase in the normal amount of soluble β1,4-galactosyltransferase or a decrease in the normal amount of membrane-bound β1,4-galactosyltransferase being indicative of an abnormal condition. The abnormal condition may be congenital dyserthropoetic anemia type II.

As discussed hereinabove, the detection may be carried out by various means including immunoassay, such as RIA or ELISA. Such formats are well known to one skilled in the art. See for example HANDBOOK OF EXPERIMENTAL IMMUNOLOGY, (D.M. Weir, Ed.) Blackwell Scientific Publications (3rd ed. 1978), which is incorporated herein by reference.

Previously isolated human cDNA covers the COOH-terminal region but lacks NH_2-terminal sequences, and therefore a cDNA clone containing the full coding region of β1,4-galactosyltransferase, including the initiation site of the membrane bound form was isolated. A gtll human placenta cDNA library was screened first with a cDNA probe then with a synthetic oligonucleotide probe Siebert and Fukuda (1986)

Proc. Natl. Acad. Sci. USA, 83, 1665-1669. Several clones were identified of which two, CT7 and J20, were characterized (see Fig.1).

Nucleotide sequencing of cDNA was accomplished by subcloning into a double stranded DNA vector which allows sequencing from both the 5' and 3' ends using synthetic oligonucleotide primers (see sequencing strategy, Fig. 1). Clone CT7 revealed a novel sequence at the 5' end while having homology to the COOH-terminal sequence of galactosyltransferase down to nucleotide 1023 suggesting that it was primed at the $(A)_8$ segment (see Fig. 2). The 5' most ATG codon (nucleotide 1 in Fig. 2) is in a consensus strong context for translation initiation (Kozak, M. (1986) Cell, 44:283-292) and is proceeded by an in-frame TAA termination codon at nucleotide-18, suggesting it could act as the translation initiation signal. A single open reading frame follows this codon, and the deduced amino acid sequence of the human β1,4-galactosyltransferase protein is 400 residues long with molecular weight of 44,111 daltons. A hydropathy plot generated from the translated sequence shows only one prominent hydrophobic segment flanked by charged amino acids on both ends, characteristic of a membrane bound domain (Fig. 3). The NH_2-terminal amino acid sequence of the soluble form of β1,4-galactosyltransferase (Appert, et al. (1986) Biochem. Biophys. Res. Comm., 138:224-229 which is incorporated herein by reference) was identified (underlined by broken line in Fig. 2).

Comparison of the coding sequence of human β1,4-galactosyltransferase to the murine and bovine sequences revealed a variation of more than an 20% (Fig. 4). Sequencing of another clone (J20) revealed that it contains a sequence beginning after the proteolytic cleavage site and continuing through the coding region to just past the stop

codon (see Fig. 1).

In a study of β1,4-galactosyltransferase expression in HeLa cells, Strous et al. found two precursor forms, 44,000 and 47,000 daltons (Strous, G.J. van Berhkof, P., Willemsen, R., Geuze, H.J., and Berger, E.G. (1985) J. Cell Biol., 97, 723-727). It is of interest that a second in-frame ATG codon exists at 37 nucleotides downstream of the putative initiation codon (Fig. 2), and it could serve as the initiation site for the lower molecular weight precursor, as proposed for the murine enzyme (Shaper, N.L., Hollis, G.F., Douglas, J.G., Kirsch, I.R., and Shaper, J.H. (1988) J. Biol. Chem., 263, 10420-10428). Both precursors were glycosylated with one N-linked oligosaccharide chain (Strous, G.J., van Berhkof, P., Willemsen, R., Geuze, H.J., and Berger, E.G. (1985) J. Cell Biol., 97, 723-727). Since N-glycosylation takes place on the lumenal sides of the ER and Golgi, evidence suggests that both precursor forms have their catalytic domain in cisternal lumen. In a steady state of cultured HeLa cells, galactosyltransferase was found to require 20 min to move from the ER to the Golgi, where it remained for an average half-life of 19 hrs (Strous, G.J., and Berger, E.G. (1982) J. Biol. Chem., 257, 7623-7628). These data suggest a mechanism in which galactosyltransferase is retarded at the level of the distal Golgi cisternae prior to release into the medium. In the HEMPAS variant cells, only membrane bound form of β1,4-galactosyltransferase is decreased (Fukuda, M.N., Masri, K.A., Dell, A., Thonar, E.J.-M, Klier, G., and Lowenthal R.M., Blood, in press). Isolation of cDNA containing the entire coding sequence for human β1,4-galactosyltransferase now allows us to use molecular genetic techniques to analyze patient cells.

The following examples are intended to illustrate but not limit the invention.

EXAMPLE I

Preparation of cDNA probe

A 982bp cDNA encoding the COOH-terminal region of human β1,4-galactosyltransferase (Appert, H.E., Rutherford, T.J., Tarr, G.E., Wiest, J.S., Thomford, N.R., and McCorquosdale, D.J. (1986) Biochem. Biophys. Res. Comm., 139, 163-168) has been inserted into the EcoRI site of pUC18 vector (Pharmacia Fine Chemicals, Piscataway, NJ). The pUC18 plasmid DNA was digested with EcoRI (Bethesda Research Institute, Bethesda, MD). The reaction was stopped by adding 0.5M EDTA to a final concentration of 15mM, then loaded on a 1% mini agarose gel. The cDNA insert band was cut out from the gel and electroeluted using an electrophoretic concentrator (Model 1750, ISCO, Lincoln, NE). The DNA was extracted once with phenol, twice with isoamyl alcohol and then precipitated with ethanol at -20°C. Labeling with [^{32}P]-dCTP using nick translation kit (Pharmacia Fine Chemicals, Piscataway, NJ) was performed at 15°C for 1 hr according to the manual provided by the supplier, then purified on mini-spin columns (Worthington Biochemicals, Freeland, NJ) with a 70-90% recovery rate.

EXAMPLE II

Preparation of oligonucleotide probe

A 21mer synthetic oligonucleotide, CTGCTTTGCCACGAGCTCCAG, which hybridizes to the sequence starting at nucleotide 40 of the 982bp, cDNA was labeled with γ-[^{32}P]-ATP (New England Nuclear, Boston, MA) using T4-kinase. Briefly, 400ng of 21mer was incubated with 10-20 units of T4 kinase and 850 μCi γ-[^{32}P]-ATP (6000 Ci/mmol) at 37°C for 1 hr. The [^{32}P]-oligonucleotide was purified on a NACS PREPAC mini column (Bethesda Research Laboratories).

EXAMPLE III
Screening of gt11 cDNA library

A gt11 human placenta cDNA library (Millan, J.L (1986) J. Biol. Chem., 261, 3112-3115) was kindly provided by Dr. J.L. Millan, at the La Jolla Cancer Research Foundation.

A total of 5 x 10^6 phage plaques on E. coli strain Y1088 lawn cells were screened. A nitrocellulose filter was placed on phage plaques for 1 minute for the first lift and 5 min for the second. The filters were soaked in 1.5 M NaCl-1M Tris, 1.5M NaCl-0.5M NaOH, and 3 x SSC for 2, 5, and 1-5 min. respectively. Filters were air dried then baked in a vacuum oven at 80°C for 2 hrs. The dried filters were prehybridized for at least one hr at 60°C in the following buffer: 5x Denhardt, 5x SET, 0.1% NaPP, 0.1% SDS, 50μg/ml herring sperm DNA. Hybridization followed at 60°C overnight in the above mentioned buffer with labeled cDNA probe (1.0 x 10^6 cpm/ml final). Filters were washed with several volumes of 2X SSC, 0.2% SDS at room temperature, then soaked with the same buffer twice at 50°C. Autoradiography was performed by exposing filters to X-OMAT AR diagnostic film (Kodak, Rochester, NY) using an intensifying screen overnight at-70°C. After 4 rounds of selection, several positive clones were obtained and further tested by probing with the 21mer synthetic oligonucleotide probe: nitrocellulose filters were soaked with prehybridization buffer (6x SSC, 1x Dendhardt's, 0.5%SDS, 0.05% naPP), containing 100μg/ml herring sperm DNA for at least 2 hrs at 50°C. Hybridization with the oligonucleotide probe was performed by soaking with the same buffer containing 20μg/ml E. coli tRNA and probe (1.0 X 10^6 cpm/ml) overnight at 50°C.(Siebert, P.D., and Fukuda, M. (1986) Proc. Natl. Acad. Sci. USA, 83, 1665-1669). Five of the clones, CT14, J18, J20, J2C, and CT7, were identified to

be positive.

EXAMPLE IV

Sequencing analysis

Phage DNA was grown on four 150x15mm LB agar plates and phage DNA was isolated according to the method of Maniatis (Maniatis, T. et al. (1982) Molecular Cloning: A laboratory Manual (Cold Spring Harbor Laboratory) Cold Spring Harbor, NY), which is incorporated herein by reference. EcoRI digestion showed that phage DNA of all 5 clones contained inserts ranging from 0.9 kb to 1.4 kb in size. DNAs were isolated from 1% mini agarose gels as described by Maniatis (Maniatis, T., et al. Supra and ligated into the dephosphorylated EcoRI site of Bluescript plasmid, (Stratagene, La Jolla, CA). Dephosphorylation was performed using bacterial alkaline phosphatase (147U/μl) (Bethesda Research Institute, Bethesda, MD) at 65°C for 1 hr. For each 200 ng of dephosphorylated vector, a three fold molar excess of insert DNA and one unit of T4 DNA ligase (Bethesda Research Institute, Bethesda, MD) were used. The reaction mixture was incubated at 15°C overnight. Transformation of XL-1 Blue competent cells was carried out according to Stratagene's provided protocol, using 1-2 ng of ligated DNA per 100μl of XL-1 Blue cells. Positive clones, identified as white colonies, were grown in liquid culture, then plasmid DNA was purified using the alkaline lysis procedure (Maniatis, T., et al., Supra and CsCl density equilibration centrifugation. Sequencing of the plasmid DNA was performed by the Sanger dideoxy chain termination procedure (Sanger, F., Nicklen, S., and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA, 74, 5463-5467) according to the Sequenase kit (United States Biochemicals, Cleveland, OH) using the dGTP labeling mix and [^{35}S]dATP (New England Nuclear, Boston, MA) as a tracer. Universal sequencing primers (KS, T3, SK, and

T7) for Bluescript plasmid, and synthetic oligonucleotides (16-17mers), were used to complete the sequencing (see Fig. 1 for sequencing strategy).

EXAMPLE V

EXPRESSION OF MEMBRANE-BOUND β1,4-GALACTOSYLTRANSFERASE

Membrane-bound β1,4 galactosyltransferase was expressed as follows: two overlapping clones, CT-7 and J20, together containing the full coding region of β1,4-galactosyltransferase, were separately cloned into bluescript plasmids (Stratagene, San Diego, CA). Both clones were NotI (Stratagene, San Diego, CA) digested, combined and ligated. Bluescript plasmid recombinants containing the full coding region of β1,4-galactosyltransferase were then isolated. The Bluescript plasmids containing the full coding region of β1,4-galactosyltransferase were then SmaI digested and religated. These Bluescript plasmids were then SmaI and Hind III (Bethesda Research Institute, Bethesda, MD) digested and ligated with similarly digested pTZ18U plasmids and recombinants were isolated. The recombinants were then EcoRI (Bethesda Research Institute, Bethesda, MD) digested and ligated with similarly digested pIN-III ompA3 plasmids (provided by Dr. Masayori Inoue, University of Medicine and Dentistry of New Jersey,) and recombinants containing the full coding region of β1,4-galactosyltransferase were isolated. The isolated pIN-III-ompA3 plasmids containing the full coding region of β1,4-galactosyltransferase were then used for expression of the β1,4-galactosyltransferase in E. coli.

E. coli was transformed by standard procedures as follows: A dry ice/ethanol bath was prepared. The cells were thawed and mixed by hand and a 100 µl aliquot placed in

a 15 ml polypropylene tube (Falcon 2059). A fresh dilution of 1.76 μl β mercaptoethanol (1:10 dil.) in high quality water was added to the 100 μl of bacteria, giving a 25 mM final concentration. The mixture was swirled and iced for 10 minutes, swirling gently every two minutes. 5 μl of plasmid DNA was added and iced for 30 minutes followed by heat pulse in a 42°C water bath for 45 seconds and iced for 2 minutes. Then 0.9 ml SOC medium was added and incubated at 37°C for 1 hour shaking at 225 rpm. Cells were plated directly, 200 μl per plate. The pellet was then resuspended in 200 μl and plated on a 100 mm plate. After autoclaving 10 mls of a 1 mg/ml tetracycline solution were added and 50 mg/ml amp. was added when temperature dropped below 55°C.

The resulting transformed E. coli produced human membrane-bound β1,4-galactosyltransferase.

EXAMPLE VI
PREPARATION OF ANTIBODIES

Antibodies specific to soluble GT were prepared as follows: 5 mg Keyhole limpet hemocyanin by (KLH) was dissolved in 0.05M phosphate buffer, pH 7.0. 7.5μL meta-maleimidobenzoyl N-hydroxysuccinimide ester (MBS) (5 mg/mL in dimethyl formamide were added and the solution incubated at room temperature for 1 hour with occasional stirring. Unbound MBS was removed by applying the solution to a G-25 column (30 cm X 0.9 cm; Pharmacia Fine Chemicals, Piscataway, NJ) and eluted with phosphate buffer, pH 7.0 containing 50 mM NaCl. Fractions were analyzed using a ultraviolet spectrophotometer (DU 20; Beckman Instruments, Brea, CA). Those exhibiting peak absorbance at 280 nm were combined and immediately mixed with 5 mg of synthetic peptide dissolved in phosphate buffer, pH 7.0. Synthetic peptides comprising the

amino acid sequence SRDKKNEPNPQRFDR (amino acids 348 through 362 in Figure 2), had been previously synthesized using an automatic peptide synthesizer (Model 430A; Applied Biosystems, Inc., Foster City, CA). The solution was incubated at room temperature for 2 hours and the reaction stopped by the addition of 1 drop of β-mercaptoethanol. The solution was applied to a Sepharose 4B column (1.8 X 33 cm; Pharmacia Fine Chemicals, Piscataway, NJ), equilibrated with 0.02 M phosphate buffer containing 0.1M NaCl. KLH containing fractions were again identified by absorbance at 280 nm. Selected fractions were stored and dialyzed against phosphate buffered saline.

A female adult New Zealand White rabbit was injected with 1 mg of peptide dissolved in 200 µl of phosphate buffered saline in Freund's Complete Adjuvant, and boosted one month later with 1 mg of peptide dissolved in 200 µl of phosphate buffered saline in Freund's Incomplete Adjuvant.

The antiserum was removed from the rabbit and passed over a column to which the bovine soluble galactosyltransferase (Sigma) was conjugated. The specific antibodies were eluted with 4M guanidine-HCl in phosphate buffered saline after washing with the phosphate buffered solution. The eluted antibodies were recovered by dialyzing the eluate against the phosphate buffered solution.

Although the invention has been described with reference to the presently-preferred embodiment, it should be understood that various modifications can be made without departing from the spirit of the invention. Accordingly, the invention is limited only by the following claims.

WE CLAIM:

1. An isolated nucleic acid sequence which encodes purified membrane-bound human β1,4-galactosyltransferase, or a functional equivalent thereof.

2. The nucleic acid sequence of claim 1 wherein the nucleic acid is selected from the group consisting of DNA, RNA, or cDNA.

3. A cDNA sequence comprising the sequence identified for membrane-bound human β1,4-galactosyltransferase in Figure 2.

4. An isolated nucleic acid sequence having the sequence identified in Figure 2 beginning with adenine at position 1 and ending with cytosine at position 1200.

5. An isolated nucleic acid sequence which encodes purified soluble human β1,4-galactosyltransferase or a functional equivalent thereof.

6. The nucleic acid sequence of claim 5 wherein the nucleic acid sequence is selected from the group consisting of DNA, RNA or cDNA.

7. The cDNA sequence of claim 5 comprising the sequence identified for soluble human β1,4-galactosyltransferase in Figure 2.

8. An isolated nucleic acid sequence having the sequence identified in Figure 2 beginning with adenine at position 231 and ending with cytosine at position 1200.

9. A vector comprising the nucleic acid sequence of either claim 1 or 5.

10. The vector of claim 9 wherein the vector is a plasmid.

11. The plasmid of claim 10 comprising pTZ18U.

12. The plasmid of claim 10 comprising pIN-III-ompA3.

13. Recombinant host cells transformed with the vector of claim 9.

14. Polypeptides produced by the recombinant host cells of claim 13.

15. Antibodies reactive with a portion of membrane-bound β1,4-galactosyltransferase identified in Figure 2 beginning with arginine corresponding to nucleotide positions 4 through 6 and ending with arginine corresponding to nucleotide positions 228 through 230.

16. Antibodies of claim 15, wherein the antibodies are monoclonal.

17. Antibodies of claim 15, wherein the antibodies are polyclonal.

18. Antibodies reactive with a portion of both soluble and membrane-bound β1,4-galactosyltransferase identified in Figure 2 beginning with threonine corresponding to nucleotide positions 231 through 233 and ending with serine corresponding to nucleotide positions 1198 through 1200.

19. Antibodies of claim 18, wherein the antibodies are monoclonal.

20. Antibodies of claim 18, wherein the antibodies are polyclonal.

21. A nucleic acid probe comprising a nucleotide sequence complementary to a portion of the nucleotide sequence between nucleotides 1 to 411 in Figure 2.

22. A method of catalyzing the transfer of galactose from UDP-galactose to acceptor sugars comprising performing the transfer in the presence of the polypeptide of claim 13.

23. A method of claim 22, wherein the acceptor sugar is N-acetylglucosamine.

24. A method of claim 22, wherein the acceptor sugar is glucose.

25. A method of diagnosing an abnormal condition in a subject comprising:

a. detecting the presence of soluble and/or membrane-bound β1,4-galactosyltransferase;

b. quantifying the relative amounts of soluble and/or membrane-bound β1,4-galactosyltransferase; and

c. comparing the amount of soluble and/or membrane-bound β1,4-galactosyltransferase to the amount in a normal subject; an increase in the normal amount of soluble β1,4-galactosyltransferase or a decrease in the normal amount of membrane-bound β1,4-galactosyltransferase being indicative of an abnormal condition.

26. The method of claim 25, wherein the abnormal condition is congenital dyserythropoietic anemia type II.

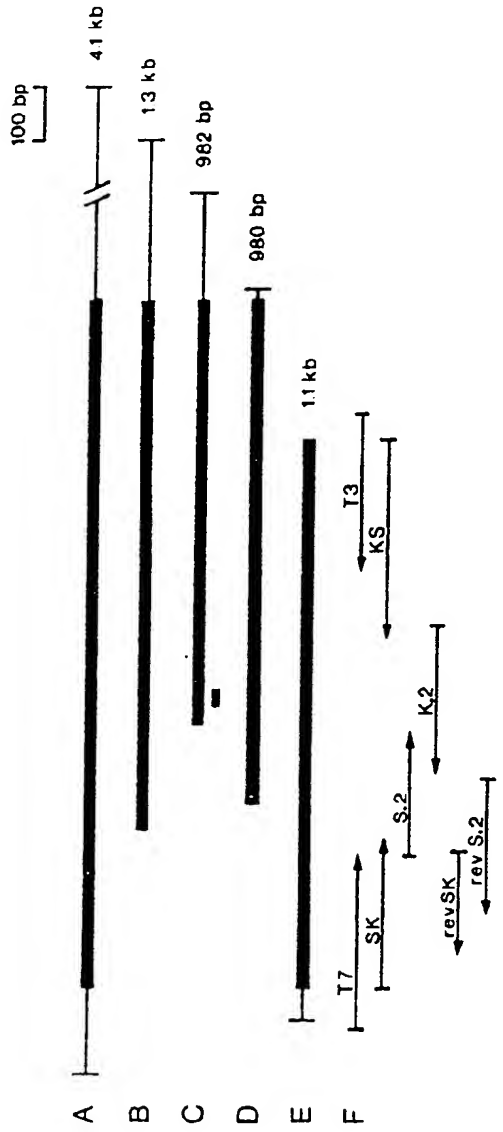

100 bp
41 kb
13 kb
982 bp
980 bp
1.1 kb
T3
KS
K.2
S.2
rev S.2
revSK
SK
T7
A
B
C
D
E
F

FIGURE 1

```
                                                                                CGGCCCGCGGGCCGGTCCGTCCCCTCCTGTAGCCCACACCCTTCTTAAAGCGGCGGCGGGAAG

                    30                                  60                                  90
ATGAGGCTTCGGGAGCCGCTCCTGAGCGGCGCCGCGATGCCAGGCGCGTCCCTACAGCGGGCCTGCCGCCTGCTCGTGGCCGTCTGCGCT
MetArgLeuArgGluProLeuLeuSerGlyAlaAlaMetProGlyAlaSerLeuGlnArgAlaCysArgLeuLeuValAlaValCysAla

                   120                                 150                                 180
CTGCACCTTGGCGTCACCCTCGTTTACTACCTGGCTGGCCGCGACCTGAGCCGCCTGCCCCAACTGGTCGGAGTCTCCACACCGCTGCAG
LeuHisLeuGlyValThrLeuValTyrTyrLeuAlaGlyArgAspLeuSerArgLeuProGlnLeuValGlyValSerThrProLeuGln

                   210                                 240                                 270
GGCGGCTCGAACAGTGCCGCCGCCATCGGGCAGTCCTCCGGGGAGCTCCGGACCGGAGGGGCCCGGCCGCCGCCTCCTCTAGGCGCCTCC
GlyGlySerAsnSerAlaAlaAlaIleGlyGlnSerSerGlyGluLeuArgThrGlyGlyAlaArgProProProProLeuGlyAlaSer

                   300                                 330                                 360
TCCCAGCCGCGCCCGGGTGGCGACTCCAGCCCAGTCGTGGATTCTGGCCCTGGCCCCGCTAGCAACTTGACCTCGGTCCCAGTGCCCCAC
SerGlnProArgProGlyGlyAspSerSerProValValAspSerGlyProGlyProAlaSerAsnLeuThrSerValProValProHis

                   390                                 420                                 450
ACCACCGCACTGTCGCTGCCCGCCTGCCCTGAGGAGTCCCCGCTGCTTGTGGGCCCCATGCTGATTGAGTTTAACATGCCTGTGGACCTG
ThrThrAlaLeuSerLeuProAlaCysProGluGluSerProLeuLeuValGlyProMetLeuIleGluPheAsnMetProValAspLeu

                   480                                 510                                 540
GAGCTCGTGGCAAAGCAGAACCCAAATGTGAAGATGGGCGGCCGCTATGCCCCCAGGGACTGCGTCTCTCCTCACAAAGTGGCCATCATC
GluLeuValAlaLysGlnAsnProAsnValLysMetGlyGlyArgTyrAlaProArgAspCysValSerProHisLysValAlaIleIle

                   570                                 600                                 630
ATTCCATTCCGCAACCGGCAGGAGCACCTCAAGTACTGGCTATATTATTTGCACCCAGTCCTGCAGCGCCAGCAGCTGGACTATGGCATC
IleProPheArgAsnArgGlnGluHisLeuLysTyrTrpLeuTyrTyrLeuHisProValLeuGlnArgGlnGlnLeuAspTyrGlyIle

                   660                                 690                                 720
TATGGCATCTATGTTATCAACCAGGCGGGAGACACTATATTCAATCGTGCTAAGCTCCTCAATGTTGGCTTTCAAGAAGCCTTGAAGGAC
TyrGlyIleTyrValIleAsnGlnAlaGlyAspThrIlePheAsnArgAlaLysLeuLeuAsnValGlyPheGlnGluAlaLeuLysAsp

                   750                                 780                                 810
TATGACTACACCTGCTTTGTGTTTAGTGACGTGGACCTCATCCCAATGAATGACCATAATGCGTACAGGTGTTTTTCACAGCCACGGCAC
TyrAspTyrThrCysPheValPheSerAspValAspLeuIleProMetAsnAspHisAsnAlaTyrArgCysPheSerGlnProArgHis

                   840                                 870                                 900
ATTTCCGTTGCAATGGATAAGTTTGGATTCAGCCTACCTTATGTTCAGTATTTTGGAGGTGTCTCTGCTCTAAGTAAACAACAGTTTCTA
IleSerValAlaMetAspLysPheGlyPheSerLeuProTyrValGlnTyrPheGlyGlyValSerAlaLeuSerLysGlnGlnPheLeu

                   930                                 960                                 990
ACCATCAATGGATTTCCTAATAATTATTGGGGCTGGGGAGGAGAAGATGATGACATTTTTAACAGATTAGTTTTTAGAGGCATGTCTATA
ThrIleAsnGlyPheProAsnAsnTyrTrpGlyTrpGlyGlyGluAspAspAspIlePheAsnArgLeuValPheArgGlyMetSerIle

                  1020                                1050                                1080
TCTCGCCCAAATGCTGTGGTCGGGAGGTGTCGCATGATCCGCCACTCAAGAGACAAAAAAAATGAACCCAATCCTCAGAGGTTTGACCGA
SerArgProAsnAlaValValGlyArgCysArgMetIleArgHisSerArgAspLysLysAsnGluProAsnProGlnArgPheAspArg

                  1110                                1140                                1170
ATTGCACACACAAAGGAGACAATGCTCTCTGATGGTTTGAACTCACTCACCTACCAGGTGCTGGATGTACAGAGATACCCATTGTATACC
IleAlaHisThrLysGluThrMetLeuSerAspGlyLeuAsnSerLeuThrTyrGlnValLeuAspValGlnArgTyrProLeuTyrThr

                  1200                                1230                                1260
CAAATCACAGTGGACATCGGGACACCGAGCTAGCGTTTTGGTACACGGATAAGAGACTGAAATTAGCCAGGGACCTCTGCTGTGTGTCTC
GlnIleThrValAspIleGlyThrProSerEnd

                  1290                                1320                                1350
TGCCAATCTGCTGGGCTGGTCCCTCTCATTTTTACCAGTCTGAGTGACAGGTCCCCTTCTTCGCTCATCATTCAGATGGCTTTCCAGATG

                  1380
ACCAGGACGAGTGGGATATTTTGCCCCCAACTTGGCTCGGCATGTGAATTC
```

FIGURE 2

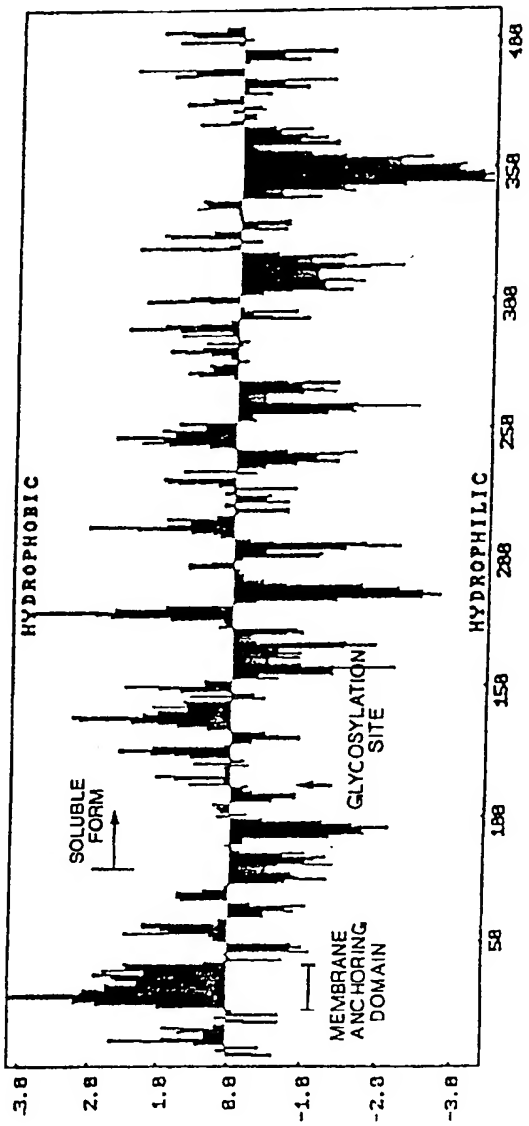

HYDROPHOBIC
HYDROPHILIC
SOLUBLE FORM
GLYCOSYLATION SITE
MEMBRANE ANCHORING DOMAIN
-3.0
-2.0
-1.0
0.0
1.0
2.0
3.0
58
108
158
208
258
308
358
408

FIGURE 3

```
                                                45
h     MRLREPLLSG*AAMPGASLQRACRLLVAVCALHLGVTLVYYLAGRDLSRLPQLVGVSTPLQGGSNSAAAIGQSSGELRTGGARPPPPLGA  90
m     --F--QF-G-S------T-----------------------S-------------ST----T-G---SK-PP--Q-PR---------V
b                                                                          -I--P-----LR-VA-----QN

                                               135
h     SSQPRPGGDSSPVVDSGPGPASNLTSVPVPHTTAL*SLPACPEESPLLVGPMLIEFNMPVDLELVAKQNPNVKMGGRYAPRDCVSPHKVA 180
m     -PK----L----GAA----LK---S-L---T--G-L-----------------D--IA-----L--K---I-T----S-K---------
b     --K--**L-AYSHPGP----G-----A---S--TR*--T-----------------I----K--QQ---L--L----T-M--I------

                                               225
h     IIIPFRNRQEHLKYWLYYLHPVLQRQQLDYGIYGIYVINQAGDTIFNRAKLLNVGFQEALKDYDYTCFVFSDVDLIPMNDHNAYRCFSQP 270
m     --------------------I----------------------M--------I----------N-----------D-R---------
b     --------------------I--------------------ESM----------K--------N---------------T-------

                                               315
h     RHISVAMDKFGFSLPYVQYFGGVSALSKQQFLTINGFPNNYWGWGGEDDDIFNRLVFRGMSISRPNAVVGRCRMIRHSRDKKNEPNPQRF 360
m     ------------------------------A----------------------HK-------------------------------
b     ----------L-------------------S-----------------Y---A-----V------I-K------------------

h     DRIAHTKETMLSDGLNSLTYQVLDVQRYPLYTQITVDIGTPS
m     ----------RF--------K-------------------R
b     --------------------M--E--------K---------
```

FIGURE 4